JOURNAL OF MULTIVARIATE ANALYSIS 32, 204–212 (1990)

A Generalized Cramér–Rao Analogue for Median-Unbiased Estimators

N. K. SUNG*

Harvard University

Communicated by the Editors

A generalized Cramér–Rao analogue for median-unbiased estimators having continuous joint density functions is given by defining a dispersion measure, *joint diffusivity*, analogous to the generalized variance in mean-unbiased estimation and extending the usual definition of median-unbiasedness to the multivariate case. The resulting inequality is valid for multivariate distributions with a vector parameter. © 1990 Academic Press, Inc.

1. INTRODUCTION

Sung, Stangenhaus, and David [3] have introduced an analogue of the Cramér–Rao inequality for median-unbiased estimators with continuous density functions depending upon a single real-valued parameter, in which the usual *mean-unbiasedness* is replaced by median-unbiasedness, variance is replaced by *diffusivity*, and the Fisher information is replaced by the absolute moment of the sample score; here, *diffusivity* is defined to be the reciprocal of twice the median-unbiased estimator's density height evaluated at its median point.

The present paper extends the Cramér–Rao analogue given by Sung *et al.* [3] to the case of multivariate observations with vector-valued parameter, with a restriction on the form of the parametric functions of interest, and compares the resulting inequality to the generalized Cramér–Rao inequality.

Let $X = (X_1, ..., X_n)$ be a random vector of n i.i.d. random variables having a joint density function $f(x; \theta)$, $\theta \in \Theta$, where x stands for the observations $x_1, ..., x_n$ and θ for the parameters $\theta_1, ..., \theta_r$. The parameter space Θ is either the Euclidean r-space R^r or a rectangle in R^r. Let $\tau_j(\theta)$ be a

Received February 20, 1989; revised May 1989.
AMS 1980 subject classification: Primary 62F10.
Key words and phrases: median-unbiased estimation; Cramér–Rao; diffusivity.

* Present address: Statistical Research Institute, Seoul National University, Seoul, Korea.

0047-259X/90 $3.00

real-valued function on Θ, which is partial-differentiable with respect to θ_i, $i=1, ..., r$, $j=1, ..., k$. Let $Y=(\delta_1(X), ..., \delta_k(X))$ be an estimator of $\tau(\theta)=(\tau_1(\theta), ..., \tau_k(\theta))$.

The generalized Cramér–Rao inequality states that, under certain regularity conditions (cf., Zacks [4, p. 194]),

$$\Sigma(Y;\theta)-C(\theta)\,I^{-1}(\theta)\,C'(\theta) \tag{1}$$

is non-negative definite for all $\theta\in\Theta$, where Y is mean-unbiased for τ, $\Sigma(Y;\theta)$ is the covariance matrix of Y, $I(\theta)=(I_{ij}(\theta))_{r\times r}$ is the information matrix with $I_{ij}(\theta) = E_\theta[(\partial/\partial\theta_i)\log f(X;\theta)\cdot(\partial/\partial\theta_j)\log f(X;\theta)]$, and $C(\theta)=(C_{ij}(\theta))_{k\times r}$ is the matrix of partial derivatives $C_{ij}(\theta)=(\partial/\partial\theta_j)\,\tau_i(\theta)$; the non-negativity of (1) implies, when $C(\theta)$ is nonsingular, that the generalized variance $\det\{\Sigma(Y;\theta)\}$ of Y is bounded below by

$$[\det\{C(\theta)\}]^2/\det\{I(\theta)\}. \tag{2}$$

The above discussion of the generalized Cramér–Rao inequality in terms of scalar observations also holds true for vector observations, as was pointed out by Anderson [1, p. 80].

In order to develop an analogue of (2) for median-unbiased estimators, as given in Section 2, a generalized definition of median-unbiasedness and an analogue of the generalized variance in mean-unbiased estimation are required.

Let $X_\alpha=(X_{\alpha 1}, ..., X_{\alpha p})$, $\alpha=1, ..., n$, be a random sample of size n from a p-variate distribution, with joint density $f(x_1, ..., x_n;\theta)$, where x_α stands for the αth observation $(x_{\alpha 1}, ..., x_{\alpha p})$, $\alpha=1, ..., n$, and $\theta=(\theta_1, ..., \theta_r)\in\Theta\subset R^r$.

Consider the problem of estimating parametric functions

$$\tau(\theta)=(\tau_1(\theta), ..., \tau_k(\theta)),$$

where $\tau_i(\theta)$ is a real-valued function on Θ. We shall denote an estimator of $\tau(\theta)$,

$$Y=(\delta_1(X_1, ..., X_n), ..., \delta_k(X_1, ..., X_n)),$$

by $Y=(\delta_1(X), ..., \delta_k(X))$, where $X=(X_1, ..., X_n)$.

For the sake of brevity, let $A=[X:\delta_1(X)<\tau_1(\theta), \delta_2(X)<\tau_2(\theta), ..., \delta_r(X)<\tau_r(\theta)]$, and let subscript(s) denote the anlogues of A with the corresponding "less than" sign(s) changed to "greater than" sign(s). For example, $A_1=[X:\delta_1(X)>\tau_1(\theta), \delta_2(X)<\tau_2(\theta), ..., \delta_r(X)<\tau_r(\theta)]$, and $A_{1,2,...,r}=[X:\delta_1(X)>\tau_1(\theta), \delta_2(X)>\tau_2(\theta), ..., \delta_r(X)>\tau_r(\theta)]$.

DEFINITION 1. $Y=(\delta_1(X), ..., \delta_k(X))$ is called *median-unbiased* for $\tau(\theta)$ if, for all $\theta \in \Theta$,

$$\begin{aligned}\Pr_\theta[A] &= \Pr_\theta[A_1] = \cdots = \Pr_\theta[A_k] \\ &= \Pr_\theta[A_{1,2}] = \cdots = \Pr_\theta[A_{k-1,k}] \\ &= \cdots \\ &= \Pr_\theta[A_{1,2,...,k}]. \end{aligned} \tag{3}$$

For the case of $p=k=r=1$, Definition 1 reduces to the usual definition of median-unbiasedness. If the joint density of Y is of continuous type, then each probability in (3) has the value $1/2^k$, and Definition 1 implies that

$$\Pr_\theta[\delta_j(X) > \tau_j(\theta)] = \Pr_\theta[\delta_j(X) < \tau_j(\theta)] = 1/2$$

for all $j=1, ..., k$ and for all $\theta \in \Theta$.

Since the criterion of median-unbiasedness cannot be met exactly for discrete distributions, we assume from now on that the joint density of Y is continuous.

We now extend the notion of diffusivity given by Sung *et al.* [3] to the case of a continuous joint density function with several parameters.

DEFINITION 2. Let $Y=(Y_1, ..., Y_k)$ be a random vector having a continuous joint density function $g(y_1, ..., y_k; \theta)$, where $\theta=(\theta_1, ..., \theta_r) \in \Theta \subset R^r$. Assume that Y is median-unbiased for $(\tau_1(\theta), ..., \tau_k(\theta))$. Then

$$1/2^k g(\tau_1(\theta), \tau_2(\theta), ..., \tau_k(\theta); \theta)$$

is the *joint diffusivity* of Y.

When $k=1$, Definition 2 reduces to the definition of diffusivity given by Sung *et al.* [3].

2. A GENERALIZED CRAMÉR–RAO ANALOGUE

In the present section we develop an analogue of (2) using extended median-unbiasedness and joint diffusivity.

We assume that $Y=(\delta_1(X), ..., \delta_r(X))$ is median-unbiased for $\tau(\theta)=(\tau_1(\theta_1), ..., \tau_r(\theta_r))$ and that Y has a continuous joint density $g(y; \theta)$, where $y=(y_1, ..., y_r)$ and $\theta=(\theta_1, ..., \theta_r) \in \Theta$. $\tau_j(\theta_j)$ is assumed to be a real-valued differentiable function of θ_j, $j=1, ..., r$.

Note that we restrict attention to the case where τ_i is a function of only

θ_i, $i = 1, 2, ..., r$, and the number of parametric functions of interest is the same as the number of unknown parameters. These two restrictions will be required in the sequel.

Suppose that the following regularity conditions hold:

$$\begin{aligned} &\text{(i)}\quad (\partial^r/\partial\theta_1 \cdots \partial\theta_r)\, f(x;\theta) \text{ exists for all } \theta \in \Theta. \\ &\text{(ii)}\quad 0 < E_\theta\, |f(X;\theta)^{-1}(\partial^r/\partial\theta_1 \cdots \partial\theta_r)\, f(X;\theta)| < \infty \text{ for all } \theta \in \Theta. \end{aligned} \tag{4}$$

THEOREM. *Under the regularity conditions* (4),

$$\frac{1}{2^r g(\tau_1(\theta_1), ..., \tau_r(\theta_r); \theta)} \geqslant |\tau_1'(\theta_1) \cdots \tau_r'(\theta_r)| \Big/ E_\theta \left| \frac{1}{f(X;\theta)} \frac{\partial^r f(X;\theta)}{\partial\theta_1 \cdots \partial\theta_r} \right|, \tag{5}$$

Note that the Cramér–Rao analogue presented by Sung *et al.* [3] is a special case of (5) in which $p = r = 1$.

Proof. The critical part of the proof is to take the rth difference of g at its median $(\tau_1(\theta_1), ..., \tau_r(\theta_r))$. For convenience, let $y = (y_1, ..., y_r)$, $dy = \prod_{j=1}^r dy_j$, $dx = \prod_{\alpha=1}^n \prod_{i=1}^p dx_{\alpha i}$, and $\theta + \Delta\theta = (\theta_1 + \Delta\theta_1, ..., \theta_r + \Delta\theta_r)$, where $\Delta\theta_j$ belongs to a neighborhood to 0, $j = 1, ..., r$.

We consider the probability:

$$\int_{\tau_r(\theta_r)}^{\tau_r(\theta_r + \Delta\theta_r)} \cdots \int_{\tau_1(\theta_1)}^{\tau_1(\theta_1 + \Delta\theta_1)} g(y; \theta + \Delta\theta)\, dy. \tag{6}$$

Note that the jth integral sign in (6) can be decomposed in two ways as

$$\int_{\tau_j(\theta_j)}^{\tau_j(\theta_j + \Delta\theta_j)} = \int_{-\infty}^{\tau_j(\theta_j + \Delta\theta_j)} - \int_{-\infty}^{\tau_j(\theta_j)}, \tag{7}$$

and

$$\int_{\tau_j(\theta_j)}^{\tau_j(\theta_j + \Delta\theta_j)} = \int_{\tau_j(\theta_j)}^{\infty} - \int_{\tau_j(\theta_j + \Delta\theta_j)}^{\infty}, \tag{8}$$

where only (7) is used initially; if we apply the decomposition given in (7) to all integral signs in (6) and use the fact that the following equality holds for all $j = 1, ..., r$, such that

$$\begin{aligned} &\int \cdots \int_{-\infty}^{\tau_j(\theta_j + \Delta\theta_j)} \cdots \int g(y; ..., \theta_j + \Delta\theta_j, ...)\, dy \\ &\quad = \int \cdots \int_{-\infty}^{\tau_j(\theta_j)} \cdots \int g(y; ..., \theta_j, ...)\, dy, \end{aligned}$$

then it can be shown that (6) is equivalent to

$$\int_A [f(x;\theta_1,...,\theta_r)-f(x;\theta_1,...,\theta_{r-1},\theta_r+\Delta\theta_r)-\cdots$$
$$-f(x;\theta_1+\Delta\theta_1,\theta_2,...,\theta_r)+\cdots$$
$$+(-1)^r f(x;\theta_1+\Delta\theta_1,...,\theta_r+\Delta\theta_r)]\,dx, \tag{9}$$

where $A=[X\colon \delta_1(X)<\tau_1(\theta),...,\delta_r(X)<\tau_r]$.

By the mean-value theorem, relation (6) can be written as

$$\prod_{j=1}^{r} [\tau_j(\theta_j+\Delta\theta_j)-\tau_j(\theta_j)]\times g(\tau_1(\theta_1)+\lambda_1[\tau_1(\theta_1+\Delta\theta_1)-\tau_1(\theta_1)],...,\tau_r(\theta_r)$$
$$+\lambda_r[\tau_r(\theta_r+\Delta\theta_r)-\tau_r(\theta_r)];\theta+\Delta\theta), \tag{10}$$

for some λ_j such that $0<\lambda_j<1$, $j=1,...,r$.

Now we divide (9) and (10) by $\prod_{j=1}^r \Delta\theta_j$ and take absolute values, moving the absolute value operation inside the integral in (9); this leads to

$$\prod_{j=1}^{r}\left|\frac{\tau_j(\theta_j+\Delta\theta_j)-\tau_j(\theta_j)}{\Delta\theta_j}\right|\times g(\tau_1(\theta_1)+\lambda_1[\tau_1(\theta_1+\Delta\theta_1)-\tau_1(\theta_1)],...,\tau_r(\theta_r)$$
$$+\lambda_r[\tau_r(\theta_r+\Delta\theta_r)-\tau_r(\theta_r)];\theta+\Delta\theta)$$
$$\leqslant\int_A\Big|[f(x;\theta_1,...,\theta_r)-f(x;\theta_1,...,\theta_{r-1},\theta_r+\Delta\theta_r)-\cdots$$
$$-f(x;\theta_1+\Delta\theta_1,\theta_2,...,\theta_r)+\cdots$$
$$+(-1)^r f(x;\theta_1+\Delta\theta_1,...,\theta_r+\Delta\theta_r)]\Big/\left(\prod_{j=1}^{r}\Delta\theta_j\right)\Big|\,dx. \tag{11}$$

As a second step, we apply the decomposition (8) to only one of the integral signs in (6), say, the first integral sign for y_r while the other integral signs are decomposed according to (7). Then we obtain the same inequality as (11) except that the region of integration is changed to $[x\colon \delta_1(x)<\tau_1(\theta_1),...,\delta_{r-1}(x)<\tau_{r-1}(\theta_{r-1}),\ \ \delta_r(x)>\tau_r(\theta_r)]$. This step generates $\binom{r}{1}$ inequalities.

As $(r-1)$ further steps, we repeat this procedure until all integral signs in (6) are decomposed by (8). Of course the last step gives an integral whose region of integration involves inequality signs that are all the same: $[x\colon\delta_1(x)>\tau_1(\theta_1),...,\delta_r(x)>\tau_r(\theta_r)]$, just as in (11).

Eventually we can express the probability (6) in 2^r ways and each way gives the same inequality as (11) except for the region of x on which the integral is evaluated. Such 2^r regions of x are disjoint and the union of all such regions of x is the appropriate Cartesian product.

Hence adding all 2^r inequalities together and letting $\Delta\theta_j \to 0$ for all $j = 1, ..., r$, we obtain

$$2^r g(\tau_1(\theta_1), ..., \tau_r(\theta_r); \theta_1, ..., \theta_r) \cdot \prod_{j=1}^{r} |\tau'(\theta_j)| \leqslant \int \left| \frac{\partial^r f(x;\theta)}{\partial\theta_1 \cdots \partial\theta_r} \right| dx. \qquad \text{Q.E.D.}$$

We identify the right-hand side of (5) as an analogue of the generalized Cramér–Rao lower bound (2) for the problem of simultaneous median-unbiased estimation of functions of a vector parameter, though the generalization has a limitation in the form of parametric functions that can be treated. An important special case, of course, is the estimation of $\theta = (\theta_1, ..., \theta_r)$ itself, which has the form required in theorem can be met.

The case $r = 2$ allows a form of the bound reminiscent of the form of the usual 2-parameter Cramér–Rao bound:

$$\frac{1}{2^2 g(\tau_1(\theta_1), \tau_2(\theta_2); \theta)} \geqslant |\tau_1'(\theta_1)\, \tau_2'(\theta_2)| \Big/ E_\theta \left| \frac{\partial^2 \log f(X;\theta)}{\partial\theta_1\,\partial\theta_2} + \frac{\partial \log f(X;\theta)}{\partial\theta_1} \frac{\partial \log f(X;\theta)}{\partial\theta_2} \right|. \tag{12}$$

In mean-unbiased estimation, the lower bound of the generalized variance for $r = 2$ is given by

$$[\tau_1'(\theta_1)\cdot\tau_2'(\theta_2)]^2 \Big/ \left(E_\theta \left[\frac{\partial \log f(x;\theta)}{\partial\theta_1} \right|^2 E_\theta \left[\frac{\partial \log f(x;\theta)}{\partial\theta_2} \right]^2 \right). - \left\{ E_\theta \left[\frac{\partial \log f(x;\theta)}{\partial\theta_1} \frac{\partial \log f(x;\theta)}{\partial\theta_2} \right] \right\}^2 \Bigg). \tag{13}$$

Comparing the denominator of (12) to that of (13), one might interpret the former as the "median-unbiased" analogue of generalized two-dimensional information.

EXAMPLE 1. Let $X_\alpha = (X_{\alpha 1}, ..., X_{\alpha p})$, $\alpha = 1, ..., n$, be observations of p characteristics on a random sample of size n from a p-variate normal distribution

$$f_{X_{\alpha 1}, ..., X_{\alpha p}}(x_{\alpha 1}, ..., x_{\alpha p}) = (1/2\pi)^{p/2} \exp\left\{ -\sum_{i=1}^{p} (x_{\alpha i} - \mu_i)^2 \Big/ 2 \right\},$$

where $E_{\mu_i} X_{\alpha i} = \mu_i$, $i = 1, ..., p$. Let $\bar{X}_i$ be the sample mean of the ith characteristic. Since $\bar{X}_i$'s are independent, $(\bar{X}_1, ..., \bar{X}_p)$ has a p-variate normal density

$$g_{\bar{X}_1, ..., \bar{X}_p}(y_1, ..., y_p) = (n/2\pi)^{p/2} \exp\left\{-(n/2) \sum_{i=1}^{p} (y_i - \mu_i)^2\right\}.$$

Obviously $(\bar{X}_1, ..., \bar{X}_p)$ is median-unbiased for $(\mu_1, ..., \mu_p)$ and has the mode at $(\mu_1, ..., \mu_p)$. Henc its joint diffusivity is given by $(2n/\pi)^{-p/2}$. Also, the right-hand side of (5) becomes $(\prod_{i=1}^{p} E_{\mu_i} |\sum_{\alpha=1}^{n} (X_{\alpha i} - \mu_i)|)^{-1}$, since $\sum_{\alpha=1}^{n} (X_{\alpha i} - \mu_i)$, $i = 1, ..., p$ are independent. But $\sum_{\alpha=1}^{n} (X_{\alpha i} - \mu_i) \sim N(0, n)$ for all i. Therefore the right-hand side has the value $(2/\sqrt{2\pi/n})^{-p} = (2n/\pi)^{-p/2}$. Hence $(\bar{X}_1, ..., \bar{X}_p)$ attains its lower bound.

EXAMPLE 2. Let $(X_1, Y_1), ..., (X_n, Y_n)$ be a sample from a bivariate normal population with parameters $\mu_1, \mu_2, \rho, \sigma_1^2, \sigma_2^2$. Assume that μ_1 and μ_2 are unknown, ρ is a fixed constant such that $|\rho| < 1$, and $\sigma_1^2 = \sigma_2^2 = 1$. $(\bar{X}, \bar{Y})$ is median-unbiased for (μ_1, μ_2), since $(\bar{X}, \bar{Y})$ is bivariate normal with parameters $\mu_1, \mu_2, \rho, 1/n, 1/n$. Its diffusivity is given by $\pi\sqrt{1-\rho^2}/2n$. It can be easily shown that the denominator of the right-hand side of (12) can be expressed as $nE\,|\rho/(1-\rho^2) - (\rho U^2 - (1+\rho^2)\,UV + \rho V^2)/(1-\rho^2)^2|$, where (U, V) is bivariate normal with parameters $0, 0, \rho, 1, 1$. A summary of computation of the left-hand side (LHS) and the right-hand side (RHS) of (12) for given values of ρ is shown below. The rightmost column indicates the ratio of LHS to RHS.

$\lvert\rho\rvert$	n(LHS)	n(RHS)	Ratio
0.0	1.57	1.57	1.00
0.1	1.56	1.53	1.02
0.2	1.54	1.44	1.07
0.3	1.50	1.32	1.14
0.4	1.44	1.16	1.24
0.5	1.36	0.99	1.37
0.6	1.26	0.81	1.56
0.7	1.12	0.61	1.83
0.8	0.94	0.41	2.29
0.9	0.68	0.21	3.32

If $\rho = 0$, $(\bar{X}, \bar{Y})$ attains its lower bound as shown in Example 1.

EXAMPLE 3. Let $X_1, ..., X_n$ be i.i.d. $N(\mu, \sigma^2)$ random variables, where both μ and σ^2 are unknown parameters. The sample mean $\bar{X}$ is a median-unbiased estimator of μ since $\bar{X} \sim N(\mu, \sigma^2/n)$. Let $S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2$. Since $S^2/\sigma^2 \sim \chi^2_{n-1}$, $S^2/\chi^2_{n-1,0.5}$ is median-unbiased for σ^2, where $\chi^2_{u,0.5}$ denotes the median of the chi-square distribution with u degrees of freedom. For convenience let $T = \bar{X}$ and $Q = S^2/\chi^2_{n-1,0.5}$. Note that Q and T are functions of the jointly minimal sufficient statistic $(\bar{X}, S^2)$ and are independent. Hence the joint diffusivity is a product of diffusivity of T and diffusivity of Q and the reciprocal of the joint diffusivity is given by

$$\sqrt{n/\pi}\,(\chi^2_{n-1,0.5})^{(n-1)/2} \exp\{-\chi^2_{n-1,0.5}/2\}/(\sigma^3 2^{n/2}\Gamma((n-1)/2)).$$

This expression can be evaluated easily since the median value of $\chi^2_{n-1,0.5}$ can be obtained from the table of the chi-square distribution. Now the denominator of the right-hand side of (12) can be expressed as

$$(2\sigma^3)^{-1} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \left(\sum_i z_i\right)\left(\sum_i z_i^2 - n - 2\right)\right| \exp\left(-\sum_i z_i^2/2\right) dz_1 \cdots dz_n,$$

where $z_i = (x_i - \mu)/\sigma$, $i = 1, 2, \ldots, n$. This integral can be evaluated approximately by use of the sample-mean Monte Carlo algorithm for small values of n. A summary of the computation is shown below. The ratio of the value of the left-hand side of the inequality to that of the right-hand side is also given.

n	σ^3(LHS)	σ^3(RHS)	Ratio
2	2.08	0.80	2.58
3	1.04	0.58	1.79
4	0.70	0.46	1.53
5	0.53	0.38	1.40
6	0.43	0.32	1.32

The median-unbiased estimator $(\bar{X}, S^2/\chi^2_{n-1,0.5})$ of (μ, σ^2) does not attain its bound. But the ratio decreases as n increases. Note that even in mean-unbiased estimation, the generalized variance of the uniformly minimum variance mean-unbiased estimator $(\bar{X}, S^2/(n-1))$ of (μ, σ^2) does not attain the Cramér–Rao lower bound, either.

ACKNOWLEDGMENTS

This paper is based on a portion of my Ph. D. dissertation written at Iowa State University. I wish to express my thanks to my advisor, H. T. David, in this connection. I am grateful to a referee for valuable suggestions that led to improved presentation.

REFERENCES

[1] ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis.* 2nd ed. Wiley, New York.

[2] SUNG, N. K. (1988). *Cramér–Rao Analogues for Median-unbiased Estimators.* Unpublished Ph. D. dissertation, Iowa State University, Ames, Iowa.

[3] SUNG, N. K., STANGENHAUS, G., AND DAVID, H. T. (1988). *A Cramér–Rao Analogue for Median-Unbiased Estimators.* Preprint Series 88-29, Department of Statistics, Iowa State University, Ames, Iowa.

[4] ZACKS, S. (1971). *The Theory of Statistical Inference.* Wiley, New York.